

PCT

(PCT Rule 61.2)

Commissioner
US Department of Commerce
United States Patent and Trademark
Office, PCT
2011 South Clark Place Room
CP2/5C24
Arlington, VA 22202
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

08 December 2000 (08.12.00)

PCT/US00/08751

640100-363

31 March 2000 (31.03.00)

01 April 1999 (01.04.99)

VAN DEN BOS, Christian et al


1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International Preliminary Examining Authority on:

25 October 2000 (25.10.00)

☐ in a notice effecting later election filed with the International Bureau on:

2. The election ☒ was

 was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

Zakaria EL KHODARY

Telephone No.: (41-22) 338.83.38



PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference 640100-363	FOR FURTHER ACTION see Notification of Transmittal of International Search Report (Form PCT ISA/220) as well as, where applicable, item 5 below.	
International application No. PCT/US 00/08751	International filing date (day/month/year) 31/03/2000	(Earliest) Priority Date (day/month/year) 01/04/1999
Applicant OSIRIS THERAPEUTICS. INC.		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 6 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the report

- a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

☐ the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

- b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

☒ contained in the international application in written form.

☒ filed together with the international application in computer readable form.

☐ furnished subsequently to this Authority in written form.

☐ furnished subsequently to this Authority in computer readable form.

☐ the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

☐ the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. ☐ **Certain claims were found unsearchable** (See Box I).

3. ☐ **Unity of invention is lacking** (see Box II).

4. With regard to the **title**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No.

☐ as suggested by the applicant.

☐ because the applicant failed to suggest a figure.

☐ because this figure better characterizes the invention.

☒ None of the figures.



FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 1 and 2 , and corresponding vector, method of detecting genes , and monoclonal antibody

2. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 3 and 4 and corresponding vector, method of detecting genes , and monoclonal antibody

3. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 5 and 6 and corresponding vector, method of detecting genes , and monoclonal antibody

4. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 7 and 8 and corresponding vector, method of detecting genes , and monoclonal antibody

5. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 9 and 10 and corresponding vector, method of detecting genes , and monoclonal antibody

6. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 11 and 12 and corresponding vector, method of detecting genes , and monoclonal antibody

7. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 13 and 14 and corresponding vector, method of detecting genes , and monoclonal antibody

8. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 00/08751

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Claims 1-24 Partially.

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

seq.id. 15 and 16 and corresponding vector, method of detecting genes , and monoclonal antibody

9. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 17 and 18 and corresponding vector, method of detecting genes , and monoclonal antibody

10. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 19 and 20 and corresponding vector, method of detecting genes , and monoclonal antibody

11. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 21 and 22 and corresponding vector, method of detecting genes , and monoclonal antibody

12. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 23 and 24 and corresponding vector, method of detecting genes , and monoclonal antibody

13. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 25 and 26 and corresponding vector, method of detecting genes , and monoclonal antibody

14. Claims: 1-24 (partially)

human mesenchymal stem cells protein with corresponding seq.id. 27,28 and 29,corresponding vector, method of detecting genes , and monoclonal antibody

Information on patent family members

PCT/US 00/08751

W0 9835022	A	13-08-1998	AU	6144498 A	26-08-1998
------------	---	------------	----	-----------	------------

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US 00/08751

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/47 C07K16/18 C12Q1/68

According to international Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

BIOSIS, EPO-Internal, MEDLINE, CHEM ABS Data, WPI Data, BIOTECHNOLOGY ABS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ROBERT STRAUSBERG: "tm56a07.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:2162100 3', mRNA sequence" EMBL DATABASE, ACCESSION NUMBER AI479234, 17 March 1999 (1999-03-17), XP002143786	1-13,17, 21-23
Y	the whole document	16
X	HILLIER L. ET AL.: "yy27e09.s1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:272488 3', mrna sequence" EMBL DATABASE, ACCESSION NUMBER N33854, 13 January 1996 (1996-01-13), XP002143787	1-13,17, 21-23
Y	the whole document	16

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

28 July 2000

Date of mailing of the international search report

08.11.00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Gurdjian, D

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 00/08751

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Location of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JANG WONHEE ET AL: "Aup1, a novel gene on mouse chromosome 6 and human chromosome 2p13." GENOMICS, vol. 36, no. 2, 1996, pages 366-368, XP000929549 ISSN: 0888-7543	14,15, 17,24
Y	abstract; figure 1 ---	16
A	WO 98 35022 A (OSIRIS THERAPEUTICS INC) 13 August 1998 (1998-08-13) abstract; figures 1-8; examples 1-6 ---	1-24
A	JAISWAL N ET AL: "Osteogenic differentiation of purified, culture-expanded human mesenchymal stem cells in vitro." JOURNAL OF CELLULAR BIOCHEMISTRY, FEB 1997, 64 (2) P295-312, XP000929561 UNITED STATES cited in the application abstract page 297, right-hand column, paragraph 3 page 298, left-hand column, paragraph 2 page 299, left-hand column, paragraph 2 -page 302, left-hand column, paragraph 1 -----	1-24

Information on patent family members

T/US 00/08751

Form PCT ISA 210 (patent family annex) (July 1992)



REC'D 12 JUL 2001

WIPO

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

14


Applicant's or agent's file reference 640100-363	FOR FURTHER ACTION See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/US00/08751	International filing date (day/month/year) 31/03/2000	Priority date (day/month/year) 01/04/1999
International Patent Classification (IPC) or national classification and IPC C07K14/00		
Applicant OSIRIS THERAPEUTICS, INC. et al		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.
2. This REPORT consists of a total of 8 sheets, including this cover sheet.
- ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☒ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☒ Certain observations on the international application

Date of submission of the demand 25/10/2000	Date of completion of this report 10.07.2001
Name and mailing address of the international preliminary examining authority:  European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Fax: +49 89 2399 - 4465	Authorized officer Marinoni, J-C Telephone No. +49 89 2399 8563





**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT**

International application No. PCT/US00/08751

I. Basis of the report

1. With regard to the **elements** of the international application (*Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17)*):

Description, pages:

1-39 as originally filed

Claims, No.:

1-24 as originally filed

Drawings, sheets:

1/5-5/5 as originally filed

Sequence listing part of the description, pages:

1-24, as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☒ contained in the international application in written form.
- ☒ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:



**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT**

International application No. PCT/US00/08751

- ☐ the description, pages:
☐ the claims, Nos.:
☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

(Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non-obvious), or to be industrially applicable have not been examined in respect of:

- ☐ the entire international application.
☒ claims Nos. 1-24 all partially.

because:

- ☐ the said international application, or the said claims Nos. relate to the following subject matter which does not require an international preliminary examination (*specify*):

- ☒ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. 14 and 15 partially are so unclear that no meaningful opinion could be formed (*specify*):
see separate sheet

- ☒ the claims, or said claims Nos. 14 and 15 partially are so inadequately supported by the description that no meaningful opinion could be formed.

- ☒ no international search report has been established for the said claims Nos. 1-24 all partially.

2. A meaningful international preliminary examination cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

- ☐ the written form has not been furnished or does not comply with the standard.
☐ the computer readable form has not been furnished or does not comply with the standard.

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement



**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT**

International application No. PCT/US00/08751

1. Statement

Novelty (N)	Yes:	Claims	4,6,16-18,23, 24
	No:	Claims	1-3, 5, 7, 8-15,19-22
Inventive step (IS)	Yes:	Claims	none
	No:	Claims	1-24
Industrial applicability (IA)	Yes:	Claims	1-24
	No:	Claims	none

2. Citations and explanations
see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:
see separate sheet



Re Item III

Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The ISA raised an objection for lack of unity of the invention under Rule 13 PCT and subsequently identified 14 groups of inventions. In the absence of payment of additional search fees, the search has been restricted to invention 1 (human mesenchymal stem cell protein with SEQ ID No.2, the nucleotide sequence encoding it having SEQ ID No.1, vectors, method of detection, monoclonal antibodies, etc...).
The following opinion is therefore restricted to invention 1 (**claims 1-24 all partially**).
2. The subject-matter of **claims 14 and 15** pertaining to undefined "active fragments, derivatives and functional analogs" covers a broad and above all completely undefined scope since
 - (i) the activity of the the protein of the invention is unknown and not even suggested in the present application; therefore the nature of "active fragments" and by extension of the claimed "agonists" (for which too no examples is described or even suggested) cannot be inferred from the content of the application,
 - (ii) proteins having the same (unknown) activity than but completely unrelated to the protein of SEQ ID No.2 are tentatively claimed (*i.e.* the functional analogs; see also the objection under **item V-3**).

Therefore, the claim is so unclear and its subject-matter not supported by the description (Article 6 PCT) that it is not sufficiently disclosed (Article 5 PCT).
Consequently, no opinion is provided concerning said subject-matter.

Re Item V

Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Reference is made to the following documents:
D1: STRAUSBERG 'tm56a07.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone



IMAGE:2162100 3', mRNA' sequence' EMBL DATABASE, ACCESSION
NUMBER AI479234, 17 March 1999

D2: HILLIER et al. 'yy27e09.s1 Soares melanocyte 2NbHM Homo sapiens cDNA
clone IMAGE:272488 3', mRNA sequence' EMBL DATABASE, ACCESSION
NUMBER N33854, 13 January 1996

D3: JANG et al. 'Aup1, a novel gene on mouse chromosome 6 and human
chromosome 2p13.' GENOMICS, Vol. 36, No. 2, 1996, pages 366-368

2. The objections and comments under **Item VIII** should be taken into consideration.
D1 discloses a DNA sequence which is 100% identical to the nucleotide sequence
of SEQ ID No. 1.

D2 discloses a DNA sequence which is 98.092% identical over 524 nt to the
nucleotide sequence of SEQ ID No. 1 (100% identical over 235 nt).

D3 discloses a DNA sequence which encodes a protein which is 90.244%
identical to the protein encoded by the nucleotide sequence of SEQ ID No. 1.
Following the objection raised under **item VIII** (especially under **item VIII-4**), it is
concluded that the subject-matter of **claim 1-3, 5, 7, 8-15 and 19-22** is not novel.

3. No inventive step can be acknowledged for the mere isolation of the full-length
cDNA encoding the protein of the invention, the sequence of the corresponding
mRNA, the sequence of the protein encoded thereby, vectors, cells, antibody
etc... It appears that claimed sequences having a certain degree of identity with
the sequences encoding the protein of SEQ ID No.2 are not exemplified, i.e.
merely expressions of desired result to obtain. Moreover, the absence of
functional limitation implies that the claimed sequences do not necessarily encode
a protein having the same biological activity than the protein of SEQ ID No. 2.

Additionally, RNAs, undefined vectors, cells, antibodies, etc... deriving from this
either non novel or non inventive protein or nucleic acid are within obvious reach
of the skilled person.

Therefore, the subject-matter of **claims 1-24** does not meet the requirements of
Article 33(3) PCT concerning inventive step.

4. In the description, no proven or even probable protein function or utility is



demonstrated (it appears that the expression of this protein is not restricted to mesenchymal cells only, see figures 3 and 4).

Therefore, the only technical problem underlying the present application that could be identified by the International Preliminary Examination Authority merely resides in the cloning of new sequences. The cloning of new sequences with no known function is not inventive *per se*.

Furthermore, Rule 5.1 a) iii) PCT requires that the description shall "disclose the invention, as claimed, in such terms that the technical problem (even if not expressly stated as such) and its solution can be understood". No inventive step can be acknowledged for claims or applications for which no solution to a clearly identified technical problem making a contribution over the prior art can be recognized. In the present case, the unfounded allegation that the expression of the protein having SEQ ID No.2 is restricted to mesenchymal cells cannot be considered as a convincing solution to a technical problem.

Therefore, **claims 1-24** do not meet the requirements of Article 33(3) PCT concerning inventive step.

Re Item VIII

Certain observations on the international application

1. In many claims, nucleotide sequences are tentatively defined by the result to be achieved (the protein they encode) and by their degree of homology to nucleic acids encoding the protein of SEQ ID No.2. The problem with this definition is that "back-translation" of the amino acid sequences generates a very large number of nucleic acid sequences (in the order of 2.5×10^{48} for a protein of 100 amino acids). It is of course possible to verify whether a given DNA sequence encodes a given protein (through translation of the DNA). It is however much more difficult to determine whether a given DNA sequence is 90% (or 95 or 98%) identical to any of the DNA sequences potentially encoding a given protein as this requires generating the whole set of "back-translations". A "normal" definition of the nucleic acid sequences as having a certain degree of homology with defined nucleic acid sequences would give the Applicant a fair protection while not introducing any unclarities. At present the lack of clarity cannot be justified as



necessary for the scope of protection: **claims 1-3** and all claims referring back to claims 1-3 are thus unnecessarily unclear.

2. Additionally, polypeptides (**claims 14 and 15**) are tentatively defined by a reference to a DNA which has a certain degree of homology to a nucleic acid encoding a polypeptide. This "circular reference" adds to the lack of clarity of the claimed subject-matter.
3. It is not clear whether the claims are intended to cover only nucleotide (resp. amino acid) sequences having 90 % (or 95, 98%, etc...) identity over the entire length of the reference sequences or also nucleotide (resp. amino acid) sequences having 90 % (or 95, 98%, etc...) identity over an undefined shorter length.
4. Furthermore, the wording of certain claims (see for example **claim 7**) is also unclear, leaving doubt as to which subject-matter is actually claimed. For example, following the comment under **item VIII-3**, it is considered that the protein of **D3** is encoded by "an isolated nucleic acid comprising at least the coding region of a human gene, said human gene containing at least a DNA sequence according to claim 1", *i.e.* by a "nucleic acid comprising a polynucleotide that is at least 90% identical to a polynucleotide encoding a polypeptide comprising an amino acid sequence of SEQ ID No.2".





INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷: C07K 14/00	A2	(11) International Publication Number: WO 00/59933 (43) International Publication Date: 12 October 2000 (12.10.00)
(21) International Application Number: PCT/US00/08751 (22) International Filing Date: 31 March 2000 (31.03.00) (30) Priority Data: 60/127,418 1 April 1999 (01.04.99) US 60/148,800 13 August 1999 (13.08.99) US (71) Applicant (for all designated States except US): OSIRIS THERAPEUTICS, INC. [US/US]; 2001 Aliceanna Street, Baltimore, MD 21231-2001 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): VAN DEN BOS, Christian [DE/US]; 2214 East Fairmount Avenue, Baltimore, MD 21231 (US). MBALAVIELE, Gabriel [CG/US]; 8488-1G Greystone Lane, Columbia, MD 21045 (US). (74) Agents: GRANT, Alan, J. et al.; Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road, Roseland, NJ 07068 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US , UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: HUMAN MESENCHYMAL DNAs AND EXPRESSION PRODUCTS (57) Abstract <p>A human mesenchymal stem cell (hMSC) cDNAs and putative polypeptides derived from Open Reading Frames contained therein are disclosed. Also disclosed are methods for utilizing the polynucleotides and polypeptides, including use as reagents for chromosomal mapping and identification, DNA fingerprinting and the possible role played by genetic mutations in the disease process, and for the generation of polyclonal and/or monoclonal antibodies specific for said polypeptides.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

HUMAN MESENCHYMAL DNAs AND EXPRESSION PRODUCTS

5

This application claims the priority of U.S. Provisional Applications 60/148,800, filed 13 August 1999, and 60/127,418, filed 1
10 April 1999, the disclosures of which are hereby incorporated by reference in their entirety.

15

BACKGROUND OF THE INVENTION

This invention relates to newly identified polynucleotide sequences corresponding to transcription products of human genes, and to complete gene sequences associated therewith and to gene expression
20 products thereof and to uses for the foregoing.

Osteoblasts, key cells in bone formation, or osteogenesis, are formed from mesenchymal stem cells. Such mesenchymal stem cells (or MSCs) of numerous mammalian species can be induced to differentiate
25 into connective tissue cell lineages by varying the *in vitro* culture conditions. Osteogenesis, the differentiation into bone cells, has been reported as a means to generate replacement bone from cultured and implanted MSCs (Bruder et al, Growth Kinetics, Self-Renewal, and the Osteogenic Potential of Purified Human Mesenchymal Stem Cells During
30 Extensive Subcultivation and Following Cryopreservation, J. Cell Biochem., 64(2):278-294 (Feb. 1997); Jaiswal et al., Osteogenic Differentiation of Purified, Culture-Expanded Human Mesenchymal Stem

Cells In Vitro, J. Cell Biochem., 64(2):295-312 (Feb. 1997), Kadiyala et al., Culture Expanded Canine Mesenchymal Stem Cells Possess Osteochondrogenic Potential In Vivo and In Vitro, Cell Transplant, 6(2):125-134 (Mar-Apr 1997)).

5

The process by which MSCs undergo osteogenic differentiation in culture is marked by the development of an osteoblastic morphology, the deposition of a hydroxyapatite mineralized extracellular matrix characteristic of osteoblasts and the presence of terminally differentiated osteocytes, as well as the expression of alkaline phosphatase (Jaiswal et al., Osteogenic Differentiation of Purified, Culture-Expanded Human Mesenchymal Stem Cells In Vitro, J. Cell Biochem., 64(2):295-312 (Feb. 1997)). Mechanisms underlying the osteogenic differentiation of human MSCs (hereafter, hMSCs) are poorly understood. Identification of proteins produced during this process would greatly facilitate the discovery and development of small molecules that target the osteoblast and its bone forming potential. Identification of these factors would be accelerated by the availability of relevant cDNA libraries constructed from hMSCs during various stages of their differentiation.

Identification and sequencing of human genes is a major goal of modern Molecular Biology. For example, by identifying genes and determining their sequences, scientists have been able to make large quantities of valuable human "gene products." These include human insulin, interferon, Factor VIII, tumor necrosis factor, human growth hormone, tissue plasminogen activator, and numerous other compounds. Additionally, knowledge of gene sequences can provide the key to treatment or cure of genetic diseases (such as muscular dystrophy and cystic fibrosis).

BRIEF SUMMARY OF THE INVENTION

5

In accordance with the present invention, Mesenchymal stem cells (MSCs) have been isolated and culture expanded from humans, and from them new cDNA libraries have been constructed from messenger ribonucleic acids (hereafter, mRNAs) isolated from hMSCs.

10

It is an object of the present invention to obtain cDNA libraries from purified and cultured MSCs and to use these isolated nucleic acids, isolated sequences, and fragments thereof, in the determination and preparation of the expression products of these nucleic acids and sequences, including fragments thereof.

15

It is a further object of the present invention to use the cDNAs so produced, and fragments thereof, as well as their expression products, as chromosomal markers for determining the location of genes within the genome, and alleles thereof, expressed during the development of differentiated mesenchymal cells.

20

It is yet another object of the present invention to provide DNA sequences for use in human "fingerprinting" whereby different individuals can be distinguished based on the sequences of the genes identified as wholly, or partly, identical to those disclosed herein.

25

It is still another object of the present invention to provide polynucleotide sequences corresponding to the genes coding for polypeptides as disclosed herein whereby such sequences can be compared with those found in similar chromosomal locations in animals, especially mammals, and most especially humans, where such animal is

30

afflicted with a disease affecting bone growth, or such other disease, or diseases, as may be affected by such genes, and thus detecting the presence of mutations in said genes leading to such diseases.

5 It is a still further object of the present invention to provide genetically engineered cells, and vectors, containing one or more copies of the nucleic acids, or DNAs, or genes, or nucleotide sequences according to the present invention, capable of expressing said peptides, or polypeptides, or proteins for rapid cloning of genes according to the
10 present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

15

Figure 1 shows the consensus sequence (SEQ ID NO: 27) for the novel DNA sequence of the invention as determined from different cDNA clones of said sequence, the latter being about 2.5 kb in length.

20

Figure 2 is a deduced amino acid sequence for the protein expressed from the sequence of Figure 1, residues 125 through 1717 and corresponding to SEQ ID NO:29. The amino acids set off between asterisks constitute a bipartite nuclear localization signal. The isoelectric point and molecular weight were also calculated for the putative protein.

25

Figure 3 shows the results of a dot blot assay for the presence of the novel DNA sequence in a variety of human tissues. For this assay, a prefabricated dot blot from Clontech (#7770-1) was hybridized using a probe generated from the 2.5 kb cDNA of Figure 1 and
30 treated according to the manufacturer's instructions. Signals due to bound probe were analyzed using a Storm 860 phosphorimager and imagequant software.

Figure 4 is a bar graph showing the distribution of the sequence of Figure 1 in a variety of human tissues based on relative mRNA abundance. The highest signal strength was in cells of adult heart and lowest was in fetal thymus. The bar graphs were generated using data from the dot blots of Figure 3 and were imported into an Excel spreadsheet. The data were then analyzed as arbitrary signal strength per tissue after subtracting background (due to non-specific hybridization). The order of the tissues in the bar graph reflects signal strength (and therefor differs from that on the dot blot of Figure 3). Figure 4(b) is a continuation of Figure 4(a).

15

DETAILED DESCRIPTION OF THE INVENTION

One aspect of the present invention is directed to nucleic acids and isolated DNA sequences and molecules, and fragments thereof (and corresponding isolated RNA sequences, and fragments thereof), including sequences complementary to the foregoing, showing sequence similarity to, or capable of hybridizing to, the DNA sequences identified in SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, and 27 or 28. The present invention is also directed to fragments or portions of such sequences which contain at least 15 bases, preferably at least 30 bases, more preferably at least 50 bases and most preferably at least 80 bases, and to those sequences which are at least 60%, preferably at least 80%, and most preferably at least 95%, especially 98%, identical thereto, and to DNA (or RNA) sequences encoding the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 29, including fragments and portions thereof and, when derived from natural sources, includes alleles thereof.

In accordance with the present invention, the term "percent identity" or "percent identical," when referring to a sequence, means that a sequence is compared to a claimed or described sequence after alignment of the sequence to be compared (the "Compared Sequence") with the described or claimed sequence (the "Reference Sequence"). The Percent Identity is then determined according to the following formula:

$$\text{Percent Identity} = 100 [1 - (C/R)]$$

10

wherein C is the number of differences between the Reference Sequence and the Compared Sequence over the length of alignment between the Reference Sequence and the Compared Sequence wherein (i) each base or amino acid in the Reference Sequence that does not have a corresponding aligned base or amino acid in the Compared Sequence and (ii) each gap in the Reference Sequence and (iii) each aligned base or amino acid in the Reference Sequence that is different from an aligned base or amino acid in the Compared Sequence, constitutes a difference; and R is the number of bases or amino acids in the Reference Sequence over the length of the alignment with the Compared Sequence with any gap created in the Reference Sequence also being counted as a base or amino acid.

20

If an alignment exists between the Compared Sequence and the Reference Sequence in which the percent identity as calculated above is about equal to or greater than a specified minimum Percent Identity then the Compared Sequence has the specified minimum percent identity to the Reference Sequence even though alignments may exist in which the hereinabove calculated Percent Identity is less than the specified Percent Identity.

30

Yet another aspect of the present invention is directed to an isolated DNA (or RNA) sequence or molecule comprising at least the

coding region of a human gene (or a DNA sequence encoding the same polypeptide as such coding region), in particular an expressed human gene, which human gene comprises a DNA sequence homologous with, or contributing to, the sequence depicted in SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27 or 28, or one at least 60%, preferably at least 80%, and most preferably at least 95%, especially 98%, identical thereto, including 100% identity, as well as fragments or portions of the coding region which encode a polypeptide having a similar function to the polypeptide encoded by said coding region. Thus, the isolated DNA (or RNA) sequence may include only the coding region of the expressed gene (or fragment or portion thereof as hereinabove indicated) or may further include all or a portion of the non-coding DNA (or RNA) of the expressed human gene.

In general, sequences homologous with and contributing to the sequences of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27 or 28 (or one at least 60%, preferably at least 80%, and most preferably at least 95% identical or homologous thereto) are from the coding region of a human gene.

20

The present invention also relates to vectors or plasmids which include such DNA (or RNA) sequences, as well as the use of the DNA (or RNA) sequences.

The sequences depicted in SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 28 are hybridizable with actual DNA and RNA sequences as derived from different human tissues. These sequences represent cDNA clones.

The sequence depicted in Figure 1 (SEQ ID NO: 27) is hybridizable with actual DNA and RNA sequences as derived from different human tissues. A number of cDNA clones have been generated. The nucleotide

sequence of Figure 1 (SEQ ID NO: 27) itself showed a nuclear location in the various tissues studied. The distribution of this sequence in various human tissues is shown in Figures 3 and 4. Some of these clones had an additional 3'-untranslated region, the presence of which is generally related to the extent to which the mRNA species remain in the cell before being turned over. See Kingman, Genetic Engineering, Blackwell, 1988, at page 313. The 3'-untranslated region may also regulate the frequency at which the mRNA is translated and thus constitute a mechanism by which the expression of the protein can be regulated. (Gray, N.K. & Wickens, M., Control of Translation Initiation in Animals, Ann. Rev. Cell Dev. Biol., 14:399-458 (1998).

The polynucleotides of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the mature polypeptide may be identical to the coding sequences present as open reading frames (ORFs) of the spolynucleotide sequences disclosed herein or may be a different coding sequence, which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature polypeptide as the polynucleotide sequences of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 28.

The polynucleotides that code for the polypeptides disclosed herein as putative proteins SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 and 29 may include, but are not limited to: only the coding sequence for the mature polypeptide; the coding sequence for the mature polypeptide and additional coding sequence such as a leader or secretory sequence, a proprotein sequence and a membrane anchor; the coding sequence for the mature polypeptide (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding

sequence 5' and/or 3' of the coding sequence for the mature polypeptide.

The polynucleotide which codes for the polypeptide of Figure 2 (SEQ ID NO:29) may include, but is not limited to: only the coding
5 sequence for the mature polypeptide; the coding sequence for the mature polypeptide and additional coding sequence such as a leader or secretory sequence, a proprotein sequence and a membrane anchor; the coding sequence for the mature polypeptide (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding
10 sequence 5' and/or 3' of the coding sequence for the mature polypeptide.

The term "polynucleotide" as used for the present invention encompasses a polynucleotide which includes only coding sequence for the
15 polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequences.

The present invention further relates to variants of the hereinabove described polynucleotides which encode fragments, analogs
20 and derivatives of the polypeptides having the amino acid sequences of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 and 29. Variants of the polynucleotide may be naturally occurring allelic variants of the polynucleotides or a non-naturally occurring variant of the polynucleotides.

25

Thus, the nucleic acids, or polynucleotides, according to the present invention may have coding sequences which are naturally occurring allelic variants of the coding sequence shown in SEQ ID NOS: 1, 3, 5, 7, 9,
11, 13, 15, 17, 19, 21, 23, 25, 27 and 28. As known in the art, an
30 allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

The present invention also includes polynucleotides, wherein the coding sequence for the mature polypeptide may be fused in the same reading frame to a polynucleotide sequence which aids in expression and secretion of a polypeptide from a host cell, for example, a leader sequence
5 which functions as a secretory sequence for controlling transport of a polypeptide from the cell and a transmembrane anchor which facilitates attachment of the polypeptide to a cellular membrane. The polypeptide having a leader sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature polypeptide. The
10 polynucleotides may also encode for a proprotein which is the mature protein plus additional 5' amino acid residues. A mature protein having a prosequence is a proprotein and is often an inactive form of the protein. Once the prosequence is cleaved an active mature protein remains.

15 Thus, for example, the polynucleotide of the present invention may encode for a mature protein, for a protein having a prosequence, for a protein having a transmembrane anchor or for a polypeptide having a prosequence, a presequence (leader sequence) and a transmembrane anchor.

20

The polynucleotides of the present invention may also have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptide of the present invention. The marker sequence may be a hexa-histidine tag supplied by a pQE-9 vector to provide
25 for purification of the mature polypeptide fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell, 37:767 (1984)).

30

Fragments of the full length polynucleotide of the present invention may be used as hybridization probes for a cDNA library to isolate

the full length cDNA and to isolate other cDNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type preferably have at least 15 bases, may have at least 30 bases and even 50 or more bases. The probe may also be used to identify a cDNA
5 clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons, and introns. An example of a screen comprises isolating the coding region of the gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence
10 complementary to that of the gene of the present invention are used to screen a library of human cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

A polynucleotide according to the present invention may have
15 at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to a polynucleotide of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 and 28 and which has an identity thereto, as hereinabove described, and which may or may not retain activity. Such polynucleotides may be employed as probes for the
20 polynucleotides or genes coding for the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 29, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

The polynucleotides according to the present invention may
25 also occur in the form of mixtures of polynucleotides hybridizable to some extent with the gene sequences containing any of the nucleotide sequences of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 and 28 including any and all fragments thereof, and which polynucleotide mixtures may be composed of any number of such polynucleotides, or fragments
30 thereof, including mixtures having at least 10, perhaps at least 30 such sequences, or fragments thereof.

Because coding regions comprise only a small portion of the human genome, identification and mapping of transcribed regions and coding regions of chromosomes is of significant interest. There is a corresponding need for reagents for identifying and marking coding
5 regions and transcribed regions of chromosomes. Furthermore, such human sequences are valuable for chromosome mapping, human identification, identification of tissue type and origin, forensic identification, and locating disease-associated genes (i.e., genes that are associated with an inherited human disease, whether through mutation,
10 deletion, or faulty gene expression) on the chromosome.

Various aspects of the present invention include each of the individual sequences, corresponding partial and complete cDNAs, genomic DNA, mRNA, antisense strands, PCR primers, coding regions, and
15 constructs. Expression vectors and polypeptide expression products, are also within the scope of the present invention, along with antibodies, especially monoclonal antibodies, to such expression products.

As used herein and except as noted otherwise, all terms are
20 defined as given below.

In accordance with the present invention, the term "gene" or "cistron" means the segment of DNA (or DNA segment) involved in producing a polypeptide chain; it includes regions preceding and following
25 the coding region (5'-and 3'- untranslated regions, or UTRs, also called leader and trailer sequences, regions, or segments) as well as intervening sequences (introns) between individual coding segments (exons), which intronic regions are typically removed during processing of post-transcriptional RNA to form the final translatable mRNA product. Of
30 course, by their nature, cDNAs contain no intronic sequences.

In accordance with the present invention, the term "DNA segment" refers to a DNA polymer, in the form of a separate fragment or as a component of a larger DNA construct, which has been derived from DNA isolated at least once in substantially pure form, i.e., free of
5 contaminating endogenous materials and in a quantity or concentration enabling identification, manipulation, and recovery of the segment and its component nucleotide sequences by standard biochemical methods, for example, using a cloning vector. Such segments are provided in the form of an open reading frame uninterrupted by internal nontranslated
10 sequences (introns), which are typically present in eukaryotic genes. Sequences of non-translated DNA may be present downstream from the open reading frame, where the same do not interfere with manipulation or expression of the coding regions.

15 The nucleic acids and polypeptide expression products disclosed according to the present invention, as well as expression vectors containing such nucleic acids, may be in "enriched form." As used herein, the term "enriched" means that the concentration of the material is at least about 2, 5, 10, 100, or 1000 times its natural
20 concentration (for example), advantageously 0.01%, by weight, preferably at least about 0.1% by weight. Enriched preparations of about 0.5%, 1%, 5%, 10%, and 20% by weight are also contemplated. The sequences, constructs, vectors, clones, and other materials comprising the present invention can advantageously be in enriched or isolated form.
25 For example, removal, via the differential display techniques described herein, of clones corresponding to ribosomal RNA and "housekeeping" genes and clones without human cDNA inserts results in a library that is "enriched" in the desired clones.

30 The DNA and RNA sequences, and polypeptides, disclosed in accordance with the present invention will commonly be in isolated form. The term "isolated" means that the material is removed from its original

environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide, or DNA, present in a living animal is not isolated, but the same polynucleotide or DNA, separated from some or all of the coexisting materials in the natural system, is
5 isolated. Such DNA could be part of a vector and/or such polynucleotide could be part of a composition, and still be isolated in that such vector or polynucleotide is not part of its natural environment.

The DNA and RNA sequences, and polypeptides, disclosed
10 in accordance with the present invention may also be in "purified" form. The term "purified" does not require absolute purity; rather, it is intended as a relative definition, and can include preparations that are highly purified or preparations that are only partially purified, as those terms are understood by those of skill in the relevant art. Individual clones isolated
15 from a cDNA library have been conventionally purified to electrophoretic homogeneity. The cDNA clones are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). By conversion of mRNA into a cDNA library, pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a
20 cDNA library from RNA and subsequently isolating individual clones from that library results in an approximately 10^6 fold purification of the native message. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.
25 Furthermore, claimed polynucleotide which has a purity of preferably 0.001%, or at least 0.01% or 0.1%; and even desirably 1% by weight or greater is expressly contemplated.

The term "coding region" refers to that portion of a human
30 gene which either naturally or normally codes for the expression product of that gene in its natural genomic environment, i.e., the region coding *in vivo* for the native expression product of the gene. The coding region can

be from a normal, mutated or altered gene, or can even be from a DNA sequence, or gene, wholly synthesized in the laboratory using methods well known to those of skill in the art of DNA synthesis.

5 In accordance with the present invention, the term "nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the proteins provided by this invention are assembled from cDNA fragments and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is
10 capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

 The term "expression product" means that polypeptide or protein that is the natural transcription product of the gene and any
15 nucleic acid sequence coding equivalents resulting from genetic code degeneracy and thus coding for the same amino acid(s).

 The term "fragment" when referring to a coding sequence means a portion of DNA comprising less than the complete human coding
20 region whose expression product retains essentially the same biological function or activity as the expression product of the complete coding region.

 When referring to a portion of a polypeptide, as used herein,
25 the terms "portion," "segment," and "fragment," refer to a continuous sequence of residues, such as amino acid residues, which sequence forms a subset of a larger sequence. For example, if a polypeptide were subjected to treatment with any of the common endopeptidases, such as trypsin or chymotrypsin, the oligopeptides resulting from such treatment would
30 represent portions, segments or fragments of the starting polypeptide. Similarly, portions, segments or fragments of polynucleotides would include those products resulting from the treatment of such polynucleotides with

endonucleases.

The term "primer" means a short nucleic acid sequence that is paired with one strand of DNA and provides a free 3'OH end at which a DNA polymerase starts synthesis of a deoxyribonucleotide chain.

The term "promoter" means a region of DNA involved in binding of RNA polymerase to initiate transcription.

The term "open reading frame (ORF)" means a series of triplets coding for amino acids without any termination codons and is a sequence (potentially) translatable into protein.

The term "exon" means any segment of an interrupted gene that is represented in the mature RNA product.

As used herein, reference to a DNA sequence includes both single stranded and double stranded DNA. Thus, the specific sequence, unless the context indicates otherwise, refers to the single strand DNA of such sequence, the duplex of such sequence with its complement (double stranded DNA) and the complement of such sequence.

In accordance with the present invention, the overall approach to identification of cDNAs from hMSCs involved measurement of gene expression during growth of human mesenchymal stem cells in culture. Cells were harvested and the total RNA content thereof was recovered. Next, using various primer combinations, reverse transcriptase and polymerase chain reaction procedures (RT-PCR) were used to produce and amplify the corresponding cDNAs, which were then screened to find regulated DNA sequences that were subsequently purified and cloned. These clones were then sequenced and used to determine a consensus sequence (one based upon the most commonly occurring bases at each

nucleotide position in a sequence after the contributing sequences are aligned by residue position). The resulting sequences were then subjected to computer database searches for novelty, and any homology with known sequences, using, for example, the BLAST program and the
5 GenBank database.

Using the RT-PCR methodology, the mRNA from the cells of interest (such as the hMSCs used in accordance with the present invention) is used to prepare a set or family of cDNAs corresponding to
10 the expressed genes of the cell. This cDNA preparation is then exhaustively hybridized with mRNA of cells not expressing the gene and resulting in removal of all sequences from the cDNA preparation that are common to the two cell samples. All of the cDNA sequences that hybridize with the other mRNA and those that remain are then hybridized
15 with mRNA from the cells expressing the gene (for example, cells from a healthy person or cells from tissues known to express the gene) to confirm that they are in fact the desired coding sequences. Because these latter clones contain sequences specific to the mRNA population of the cells of interest, they can subsequently be amplified and characterized
20 using further rounds of PCR and the general techniques of molecular biology.

In accordance with the foregoing, a cDNA library was generated and corresponds to the sequences of SEQ ID NOS: 1, 3, 5, 7,
25 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 and 28. Probes based on these cDNAs can be used to identify the relevant transcripts, using Northern Blotting Analysis methods well known in the art to localize these sequences within cells of various tissues. For example, the heaviest distribution of the gene coding for the polypeptide of Figure 2 (SEQ ID
30 NO: 29) was in heart tissue, as shown in Figures 3 and 4.

In accordance with the present invention, cDNA was quantified by spotting 0.5 μ l aliquots of standards and samples on ethidium agarose plates prepared as suggested in the instructions from the manufacturer (Stratagene, La Jolla, CA). Plates were incubated at room temperature for 15 minutes and DNA was visualized by UV transillumination. The respective cDNAs were then quantified by comparing spot intensities of the samples with those of the standards (the latter consisting of appropriate dilutions of 1 kb ladders (from Life Technology)).

10

Aliquots of each amplified library were excised and plasmids from randomly chosen colonies were analyzed by restriction nuclease analysis. In accordance with the present invention, plasmid DNA was digested with both EcoRI and XhoI nucleases (New England Biolabs) and the resulting restriction fragments were separated on 1.5% agarose gel electrophoresis. The cDNA inserts ranged in size from less than 1 kbp to larger than 4 kbp (where 1 kbp = 1,000 nucleotide base pairs of duplex DNA).

20

Each of the DNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. The sequences can be used as diagnostic probes for the presence of a specific mRNA in a particular cell type as well as in genetic linkage analysis (polymorphisms). Further, the sequences can be used as probes for locating gene regions associated with genetic disease.

The nucleotide and gene sequences of the present invention are also valuable for chromosome identification. Each sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. The mapping of the

30

polynucleotides to specific chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

5 Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-30 bp) from the sequences disclosed herein. Computer analysis of these sequences is used to rapidly select primers that do not span more than one exon in the corresponding genomic DNA, which would otherwise complicate the amplification
10 process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the sequences or subsequences disclosed herein will yield an amplified fragment.

15 PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more clones can be assigned per day using a single thermal cycler, as is well known in the art. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of
20 fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map a sequence, or part of a sequence, to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome
25 specific-cDNA libraries.

Fluorescence in situ hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with
30 cDNA as short as 500 or 600 bases; however, clones larger than 2,000 bp have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. FISH requires use of

the clone from which the sequence was derived, and the longer the better. For example, 2,000 bp is good, 4,000 is better, but more than 4,000 is probably not necessary to get good results a reasonable percentage of the time. For a review of this technique, see Verma et al.,
5 Human Chromosomes: a Manual of Basic Techniques. Pergamon Press, New York (1988).

Reagents for chromosome mapping can be used individually (to mark a single chromosome or a single site on that chromosome) or as
10 panels of reagents (for marking multiple sites and/or multiple chromosomes). Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

15 Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man
20 (available on line through Johns Hopkins University Welch Medical Library)). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically close genes).

25 Next, it is necessary to determine if there are differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

30 With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region

associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb.)

5 Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that cDNA sequence. Ultimately, complete sequencing of genes from several individuals is required to
10 confirm the presence of a mutation and to distinguish mutations from polymorphisms.

 In addition to the foregoing, the sequences of the invention, as broadly described, can be used to control gene expression through
15 triple helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al, Nucl. Acids Res.,
20 6:3073 (1979); Cooney et al, Science, 241:456 (1988) ; and Dervan et al, Science, 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem., 56:560 (1991) ; Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from
25 DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide. Antisense RNA or oligonucleotide hybridization may
30 also lead to RNase H activation and hence destruction of the molecules involved in the hybrid.

The present invention is also a useful tool in gene therapy, which requires isolation of the disease-associated gene in question as a prerequisite to the insertion of a normal gene into an organism to correct a genetic defect. The high specificity of the cDNA probes according to
5 this invention have promise of targeting such gene locations in a highly accurate manner.

The sequences of the present invention, as broadly defined, and including subsequences and fragments thereof, are also useful for
10 identification of individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands
15 for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP.

20 However, RFLP is a pattern based technique, which does not require the DNA sequence of the individual to be sequenced. Portions of the sequences of the present invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. These
25 sequences can also be used to prepare PCR primers for amplifying and isolating such selected DNA. One can, for example, take part of the sequence of the invention and prepare two PCR primers from the 5' and 3' ends of the sequence, or fragment of the sequence. These are used to amplify an individual's DNA, corresponding to the sequence. The
30 amplified DNA is sequenced.

Panels of corresponding DNA sequences from individuals, made this way, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences, due to allelic differences. The sequences of the present invention can be used to
5 particular advantage to obtain such identification sequences from individuals and from tissue. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases.
10 Each of the fragments or complete coding sequences comprising a part of the present invention can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate
15 individuals.

If a panel of reagents from the sequences according to the present invention is used to generate a unique ID database for an individual, those same reagents can later be used to identify tissue from
20 that individual. Positive identification of that individual, living or dead can be made from extremely small tissue samples.

Another use for DNA-based identification techniques is in forensic biology. PCR technology can be used to amplify DNA sequences
25 taken from very small biological samples. In one prior art technique, gene sequences are amplified at specific loci known to contain a large number of allelic variations, for example the DQ α class II HLA gene (Erlich, H., PCR Technology, Freeman and Co. (1992)). Once this specific area of the genome is amplified, it is digested with one or more restriction
30 enzymes to yield an identifying set of bands on a Southern blot probed with DNA corresponding to the DQ α class II HLA gene. In accordance with the present invention, it is clear from the results depicted in Figure 3

and 4 that the novel gene signal according to the present invention is found in many different tissues of the body.

5 The sequences of the present invention can be used to provide polynucleotide reagents specifically targeted to additional loci in the human genome, and can enhance the reliability of DNA-based forensic identifications. Those sequences targeted to noncoding regions are particularly appropriate. As mentioned above, actual base sequence information can be used for identification as an accurate alternative to
10 patterns formed by restriction enzyme generated fragments. Reagents for obtaining such sequence information are within the scope of the present invention. Such reagents can comprise complete genes, parts of genes or corresponding coding regions, or fragments of at least 15 bp, preferably at least 18 bp.

15

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue
20 prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar manner, these reagents can be used to screen tissue cultures for contamination.

25

Sequences that match perfectly to several different genes can be detected by hybridizing to chromosomes: if many chromosomal loci are observed, the sequence (or a close variant) is in more than one gene. This problem can be circumvented by using the 3'-untranslated part of the cDNA alone as a probe for the chromosomal location or for the
30 full-length cDNA or gene. The 3'-untranslated region is more likely to be unique within gene families, since there is no evolutionary pressure to conserve a coding function of this region of the mRNA.

The cDNA libraries disclosed according to the present invention ideally use directional cloning methods so that either the 5' end of the cDNA (likely to contain coding sequence) or the 3' end (likely to be a non-coding sequence) can be selectively obtained.

Using the sequence information provided herein, the polynucleotides of the present invention can be derived from natural sources or synthesized using known methods. The sequences falling within the scope of the present invention are not limited to the specific sequences described, but include human allelic and species variations thereof. Allelic variations can be routinely determined by comparison of one sequence with a sequence from another individual of the same species. Furthermore, to accommodate codon variability, the invention includes sequences coding for the same amino acid sequences as do the specific sequences disclosed herein. In other words, in a coding region, substitution of one codon for another which encodes the same amino acid is expressly contemplated. (Coding regions can be determined through routine sequence analysis.)

20

In a cDNA library there are many species of mRNA represented. Each cDNA clone can be interesting in its own right, but must be isolated from the library before further experimentation can be completed. In order to sequence any specific cDNA, it must be removed and separated (i.e. isolated and purified) from all the other sequences. This can be accomplished by many techniques known to those of skill in the art. These procedures normally involve identification of a bacterial colony containing the cDNA of interest and further amplification of that bacteria. Once a cDNA is separated from the mixed clone library, it can be used as a template for further procedures such as nucleotide sequencing.

30

The present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

15

Thus, the present invention is not restricted to such constructs or sequences alone but also includes expression vehicles, which may include plasmids, viruses, or any other expression vectors, including cells and liposomes, containing any of the nucleic acids, nucleotide sequences, DNAs, RNAs, or fragments thereof, as disclosed according to the present invention. Furthermore, this will be true regardless of whether such sequences are coding sequences or non-coding sequences and whether such coding sequences code for all or part of the expression products as disclosed herein, so long as such expression products, or fragments thereof, exhibit some utility in keeping with the invention disclosed herein. Thus, while the present invention includes an isolated DNA sequence, or nucleic acid, that expresses a human protein when in a suitable expression system, for example, a cell-free, or *in vitro*, expression system, such system may also be contained in, or part of, a suitable expression vehicle, or vector, be that a cell, a plasmid, a virus, or other operative expression vector.

30

Such expression systems, especially where part of an expression vehicle, will commonly require some promoter region that may include a promoter different from that normally associated *in vivo* with the genes coding for the gene expression products and proteins disclosed according to the present invention. Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include *lacI*, *lacZ*, T3, T7, *gpt*, λ *P_o*, and *trc*. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described construct(s). The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a procaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, L., Basic Methods in Molecular Biology, 1986) .

The constructs in host cells can be used in a conventional manner to produce the gene product coded by the recombinant sequence. Alternatively, the encoded polypeptide, once the sequence is known from the cDNAs, or from isolation of the pure product, can be synthetically produced by conventional methods of peptide synthesis, either manual or automated.

Thus, in accordance with the present invention, once the coding sequence is known, or the gene is cloned which encodes the polypeptide, conventional techniques in molecular biology can be used to

obtain the polypeptide. More generally, the present invention includes all polypeptides coded for by any and each of the DNA or RNA sequences disclosed herein, including fragments of said polypeptides, as well as derivatives and functional analogs thereof.

5

At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. (Fragments are useful, for example, in generating antibodies
10 against the native polypeptide.)

Alternatively, the DNA encoding the desired polypeptide can be inserted into a host organism and expressed. The organism can be a bacterium, yeast, cell line, or multicellular plant or animal. The literature
15 is replete with examples of suitable host organisms and expression techniques. For example, polynucleotide (DNA or mRNA) can be injected directly into muscle tissue of mammals, where it is expressed. This methodology can be used to deliver the polypeptide to the animal, or to generate an immune response against a foreign polypeptide. Wolff, et al.,
20 Science, 247:1465 (1990); Felgner, et al., Nature, 349:351 (1991). Alternatively, the coding sequence, together with appropriate regulatory regions (i.e., a construct), can be inserted into a vector, which is then used to transfect a cell. The cell (which may or may not be part of a larger organism) then expresses the polypeptide.

25

The present invention further relates to polypeptides having an amino acid sequence selected from SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 29, as well as fragments, analogs and derivatives of such polypeptide.

30

The terms "fragment," "derivative" and "analog," when referring to the polypeptides disclosed herein also mean polypeptides that

retain essentially the same biological function or activity as said polypeptides. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide. Such fragments, derivatives and analogs must have sufficient
5 similarity to the polypeptides disclosed herein so that activity of the native polypeptide is retained.

The polypeptides of the present invention may be recombinant polypeptides, natural polypeptides or synthetic polypeptides, preferably
10 recombinant polypeptides.

"Recombinant," as used herein, means that a protein is derived from recombinant (e.g., microbial or mammalian) expression systems. "Microbial" refers to recombinant proteins made in bacterial or
15 fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Protein expressed in most bacterial cultures, e.g., E. coli, will be free of glycosylation modifications; protein expressed in yeast will have a
20 glycosylation pattern different from that expressed in mammalian cells.

The fragment, derivative or analog of a polypeptide of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 and 29 may be (i) one in which one or more of the amino acid residues are substituted with a
25 conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a
30 compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or

a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the abilities of those skilled in the art in view of the teachings herein.

5

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity. When applied to polypeptides, the term "isolated" has its already stated meaning.

10

The polypeptides of the present invention include the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 29, in particular the mature polypeptide, as well as polypeptides which have at least 70% identity to these polypeptides, or which have, at
15 least 90% identity to these polypeptides, still more preferably at least 95% identity to these polypeptides and also include portions of such polypeptides with such portion generally containing at least 30 amino acids and more preferably at least 50 amino acids.

20

Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may
25 be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides
30 of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or

transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector, either of which may be in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media
5 modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

10 The polynucleotides of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40;
15 bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

20 In accordance with the present invention, an appropriate DNA sequence or segment may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into the appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in
25 the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (for example, a promoter sequence) to direct mRNA synthesis. As representative examples
30 of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli*. *lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their

viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

5 In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in E. coli.

10

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

15

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as E. coli, Streptomyces, Salmonella typhimurium; fungal cells, such as yeast; insect cells such as Drosophila S2 and Spodoptera Sf9; animal cells such as CHO, COS or Bowes melanoma; 20 adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

25 "Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is 30 transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably

include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous protein upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by *Sambrook, et al., Molecular Cloning: A Laboratory Manual*, Second Edition, (Cold Spring Harbor, N.Y., 1989), Wu et al, *Methods in Gene Biotechnology* (CRC Press, New York, NY, 1997), and *Recombinant Gene Expression Protocols*, in *Methods in Molecular Biology*, Vol. 62, (Tuan, ed., Humana Press, Totowa, NJ, 1997), the disclosures of which are hereby incorporated by reference.

Transcription of the DNA encoding the polypeptides according to the present invention by higher eukarotes can be increased by insertion of an enhancer sequence into the vector. Such enhancers have been known for some time and are usually cis-acting elements of DNA, usually anywhere from 10 to 300 bp that act on a promoter to

increase transcription. Common examples include the SV40 enhancer, the cytomegalovirus early promoter enhancer, the polyoma enhancer and the enhancers found in adenovirus.

5 Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can
10 be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing
15 secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

20 Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will
25 comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and
30 *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

10

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is derepressed by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

15

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

20
25
30

Recombinant protein produced in bacterial culture is conveniently isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary,
5 in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

10

The protein, its fragments or other derivatives, or analogs thereof, or cells expressing them, can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal, monoclonal, chimeric, single chain, Fab fragments, or the
15 product of an Fab expression library. Various procedures known in the art may be used for the production of polyclonal antibodies.

Antibodies generated against the polypeptide corresponding to a sequence of the present invention can be obtained by direct injection
20 of the polypeptide into an animal or by administering the polypeptide to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies binding the whole native polypeptide. Such antibodies can then be used
25 to isolate the polypeptide from tissue expressing that polypeptide. Moreover, a panel of such antibodies, specific to a large number of polypeptides, can be used to identify and differentiate such tissue.

For preparation of monoclonal antibodies, any technique
30 which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell

hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

5

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention.

10

The antibodies can be used in methods relating to the localization and activity of the protein sequences of the invention, e.g., for imaging these proteins, measuring levels thereof in appropriate physiological samples and the like.

15

In carrying out the procedures of the present invention it is of course to be understood that reference to particular buffers, media, reagents, cells, culture conditions and the like are not intended to be limiting, but are to be read so as to include all related materials that one of ordinary skill in the art would recognize as being of interest or value in the particular context in which that discussion is presented. For example, it is often possible to substitute one buffer system or culture medium for another and still achieve similar, if not identical, results. Those of skill in the art will have sufficient knowledge of such systems and methodologies so as to be able, without undue experimentation, to make such substitutions as will optimally serve their purposes in using the methods and procedures disclosed herein.

20

25

Specific embodiments of the invention will now be further described in more detail in the following non-limiting examples and it will be appreciated that additional and different embodiments of the teachings of the present invention will doubtless suggest themselves to those of

30

skill in the art and such other embodiments are considered to have been inferred from the disclosure herein.

5

EXAMPLE

The proteins encoded by the nucleotide sequences of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 and 28 are expressed in U2OS cells. This is achieved by selectively PCR amplifying the coding regions thereof (based on the available open reading frames) and then cloning the resulting amplicon into a suitable mammalian expression vector. One such vector is pcDNA3.1 (sold by Invitrogen - #K4800-01). The expression of the protein encoded by the described polynucleotide sequence is detected in either of two ways: by use of specific antibodies raised against peptides derived from the amino acid sequence or by use of antibodies against tags added during the cloning procedure. Examples of such tags are the V5 epitope or a poly-histidine sequence as contained in the pcDNA3.1 vector. In order to accomplish this, cells will normally be transfected with the expression construct and cultured for 1 to 5 days. Cells will then be lysed and their protein content analyzed by western blotting using the above antibodies as appropriate. Cells will also be analyzed for the subcellular localization of the protein encoded by the described polynucleotide sequence by transfecting cells in suitable chambers, culturing them for 1 to 5 days and fixing them *in situ*. Such cells will then be analyzed for the presence and localization of the encoded protein by staining cells with the above-referenced antibodies. Alternatively, cells will be transfected with an expression system in which the protein encoded by the described polynucleotide sequence is fused to a directly detectable tag such as green fluorescent protein (GFP). The expression and localization of the protein encoded by the described polynucleotide sequence is then detected by analyzing that of GFP.

For purposes of identification of the polypeptides disclosed herein, each such polypeptide is listed in the table below along with its calculated molecular weight (Daltons) and its expected isoelectric point (pI).

5 Table 1.

	<u>SEQ ID NO:</u>	<u># Residues</u>	<u>Mol. Wt.</u>	<u>pI</u>
	2	410	45786.9	8.96
10	4	227	26152.3	8.48
	6	275	30781.6	10.00
	10	84	8913.2	9.35
	12	281	30386.7	9.35
	14	322	32977.3	9.27
15	16	141	16444.4	9.34
	18	219	24418.4	9.07
	22	56	6356.3	7.85
	24	344	37375.6	5.82
	26	208	23864.9	9.71
20	29	531	60,576.6	9.63

The polypeptides of SEQ ID NOS: 8 and 20 corresponded only to partial sequences and thus no values could be calculated and such sequences are not in the table.

25

All of the polynucleotides from which these polypeptide sequences are derived are cDNAs isolated during a differential screen of osteogenic mesenchymal stem cells (MSCs) cultured for 4 days in the presence of osteogenic supplements.

30

What Is Claimed Is:

1. An isolated nucleic acid comprising a polynucleotide that is at least 90% identical to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 29.
2. An isolated nucleic acid comprising a polynucleotide that is at least 95% identical to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 29.
3. An isolated nucleic acid comprising a polynucleotide that is at least 98% identical to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 29.
4. An isolated nucleic acid comprising RNA corresponding to any of the DNA sequences or fragments of claims 1, 2 or 3.
5. An isolated nucleic acid comprising a DNA sequence identical to a sequence selected from the group consisting of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 and 28 and the complements of these.
6. An isolated nucleic acid comprising RNA corresponding to the DNA sequence of Claim 5.

7. An isolated nucleic acid comprising at least the polypeptide coding region of a human gene, said human gene containing a DNA sequence according to Claim 1.
- 5 8. An isolated nucleic acid comprising at least the polypeptide coding region of a human gene which contains the DNA sequence of Claim 5.
9. The isolated nucleic acid of claim 8 which expresses a
10 human protein when in a suitable expression system.
10. A vector comprising the DNA sequence of claim 1.
11. A vector comprising the DNA sequence of claim 3.
- 15 12. A vector comprising the DNA sequence of claim 5.
13. A vector comprising the DNA sequence of claim 9.
- 20 14. A polypeptide coded for by the DNA sequence of claim 7 and active fragments, derivatives and functional analogs thereof.
15. A polypeptide coded for by the DNA sequence of claim 8 and active fragments, derivatives and functional analogs thereof.
- 25 16. A polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 29.
- 30 17. A genetically engineered cell having inserted into the genome thereof the DNA of Claim 7.

18. A process for producing cells for expressing a polypeptide using genetically engineered cells of claim 17.

19. An isolated DNA sequence comprising a fragment of a DNA
5 of claim 5, wherein said fragment comprises at least 15 sequential bases of said sequence.

20. An isolated DNA sequence comprising a fragment of DNA of
10 claim 5, wherein said fragment comprises at least 30 sequential bases of said sequence.

21. An isolated DNA sequence comprising a fragment of DNA of
claim 5, wherein said fragment comprises at least 50 sequential bases of
said sequence.

15

22. An isolated DNA sequence comprising a fragment of DNA of
claim 5, wherein said fragment comprises at least 80 sequential bases of
said sequence.

20 23. A method of detecting genes within the human genome comprising contacting a sample of said genome with an isolated DNA selected from the group consisting of the DNAs of claims 19, 20, 21, and 22.

25 24. A monoclonal antibody against a polypeptide selected from the group consisting of the polypeptides of claims 14 and 15.

30

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 October 2000 (12.10.2000)

PCT

(10) International Publication Number
WO 00/59933 A3

(51) International Patent Classification: C12N 15/12,
C07K 14/47, 16/18, C12Q 1/68

(21) International Application Number: PCT/US00/08751

(22) International Filing Date: 31 March 2000 (31.03.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/127,418 1 April 1999 (01.04.1999) US
60/148,800 13 August 1999 (13.08.1999) US

(71) Applicant (for all designated States except US): OSIRIS
THERAPEUTICS, INC. [US/US]; 2001 Aliceanna
Street, Baltimore, MD 21231-2001 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): VAN DEN BOS,
Christian [DE/US]; 2214 East Fairmount Avenue, Bal-
timore, MD 21231 (US). MBALAVIELE, Gabriel
[CG/US]; 8488-1G Greystone Lane, Columbia, MD 21045
(US).

(74) Agents: GRANT, Alan, J. et al.; Carella, Byrne, Bain,
Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road,
Roseland, NJ 07068 (US).

(81) Designated States (national): AL, AM, AT, AU, AZ, BA,
BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM,
EE, ES, FI, GB, GD, GE, GH, GM, HU, ID, IL, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD,
SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent
(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent
(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU,
MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— With international search report.

(88) Date of publication of the international search report:
15 February 2001

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: HUMAN MESENCHYMAL DNAs AND EXPRESSION PRODUCTS

(57) Abstract: A human mesenchymal stem cell (hMSC) cDNAs and putative polypeptides derived from Open Reading Frames contained therein are disclosed. Also disclosed are methods for utilizing the polynucleotides and polypeptides, including use as reagents for chromosomal mapping and identification, DNA fingerprinting and the possible role played by genetic mutations in the disease process, and for the generation of polyclonal and/or monoclonal antibodies specific for said polypeptides.

WO 00/59933 A3



